NeoBiota 18: 173–191 (2013)
doi: 10.3897/neobiota.18.4016
www.pensoft.net/journals/neobiota

RESEARCH ARTICLE

A peer-reviewed open-access journal
NeoBiota
Advancing research on alien species and biological invasions

Representing uncertainty in a spatial invasion model that incorporates human-mediated dispersal

Frank H. Koch[1], Denys Yemshanov[2], Robert A. Haack[3]

1 *USDA Forest Service, Southern Research Station, Eastern Forest Environmental Threat Assessment Center, 3041 Cornwallis Road, Research Triangle Park, NC 27709 USA* **2** *Natural Resources Canada, Canadian Forest Service, Great Lakes Forestry Centre, 1219 Queen Street East, Sault Ste. Marie, ON, P6A 2E5, Canada* **3** *USDA Forest Service, Northern Research Station, 1407 S. Harrison Road, East Lansing, MI 48823 USA*

Corresponding author: *Frank H. Koch* (fhkoch@fs.fed.us)

Academic editor: *S. Worner* | Received 20 September 2012 | Accepted 5 March 2013 | Published 13 September 2013

Citation: Koch FH, Yemshanov D, Haack RA (2013) Representing uncertainty in a spatial invasion model that incorporates human-mediated dispersal. In: Kriticos DJ, Venette RC (Eds) Advancing risk assessment models to address climate change, economics and uncertainty. NeoBiota 18: 173–191. doi: 10.3897/neobiota.18.4016

Abstract

Most modes of human-mediated dispersal of invasive species are directional and vector-based. Classical spatial spread models usually depend on probabilistic dispersal kernels that emphasize distance over direction and have limited ability to depict rare but influential long-distance dispersal events. These aspects are problematic if such models are used to estimate invasion risk. Alternatively, a geographic network model may be better at estimating the typically low likelihoods associated with human-mediated dispersal events, but it should also provide a reasonable account of uncertainties that could affect perception of its risk estimates. We developed a network model that assesses the likelihood of dispersal of invasive forest pests in camper-transported firewood in North America. We built the model using data from the U.S. National Recreation Reservation Service, which document visitor travel between populated places and federal campgrounds across the U.S. and Canada. The study area is depicted as a set of coarse-resolution map units. Based on repeated simulations, the model estimates the probability that each unit is a possible origin and destination for firewood-facilitated forest pest invasions. We generated output maps that summarise, for each U.S. state and Canadian province, where (outside the state or province) a camper-transported forest pest likely originated. Treating these output maps as a set of baseline scenarios, we explored the sensitivity of these "origin risk" estimates to additive and multiplicative errors in the probabilities of pest transmission between locations, as well as random changes in the structure of the underlying travel network. We found the patterns of change in the origin risk estimates due to these alterations to be consistent across all states and provinces. This indicates that the network model behaves predictably in the presence of uncertainties, allowing future work to focus on closing knowledge gaps or more sophisticated treatments of the impact of uncertainty on model outputs.

Copyright *Frank H. Koch et al.* This is an open access article distributed under the terms of the Creative Commons Attribution License 3.0 (CC-BY), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Keywords
Human-mediated dispersal, firewood, forest pests, pest risk mapping, network modelling, uncertainty

Introduction

The spread of invasive alien species is largely facilitated by human activities such as trade and travel (Sakai et al. 2001; Horan et al. 2002; Hulme et al. 2008). Thus, reliable estimates of spread rates and patterns for these species depend upon the adequate representation of human movement in particular social and economic contexts (Keller et al. 2008; Wilson et al. 2009). Fundamentally, most modes of human-mediated dispersal are directional and vector-based, in the sense that they involve species transport by some physical agent, activity, or mechanism along defined routes connecting discrete locations (Carlton and Ruiz 2003). Classical two-dimensional spatial spread models usually depict spread as a diffusion or travelling wave process involving one or more dispersal kernels (Andow et al. 1990; Hastings 1996; Sharov and Liebhold 1998; Lewis and Pacala 2000). In these kernels, the probability of dispersal from one location to another is a function of the distance, rather than the direction, between them (Nathan 2006). Furthermore, while fat-tailed dispersal kernels (i.e., kernels where the dispersal probability declines slowly with distance) may provide more accurate depiction of long-distance dispersal, they are difficult to fit empirically because data describing long-distance dispersal events are typically rare (Nathan et al. 2003; Hastings et al. 2005). These two aspects – an emphasis on distance over direction and the challenge of adequately characterizing long-distance dispersal – are noteworthy limitations when classical spread models are subsequently used to assess invasion risks (e.g., to forecast which locations in an area of interest are most likely to be invaded within a specified time horizon).

However, there has been increased recognition of the utility of network-based modelling approaches to depict human-mediated dispersal of invasive species (Hulme 2009). Such models, which describe a species' movement via vectors, or links, between a set of interconnected nodes, have been perhaps most commonly applied for invaders of marine environments (e.g., Floerl et al. 2009), but also to depict the movement of invasive organisms through national- and global-scale commercial trade networks (Harwood et al. 2010; Kaluza et al. 2010). A key feature of the network-based approach is that the physical distance between nodes is far less important than their level of connectivity (Moslonka-Lefebvre et al. 2011); basically, the amount of movement along a vector between two nodes replaces the vector's length as the principal, or sometimes only, determinant of the likelihood of spread. By downplaying distance in favor of connectivity, network-based models may be better suited than classical spread models for depicting long-distance dispersal events (e.g., the transcontinental movement of an organism via shipped cargo), as long as the events occur within the context of the network's underlying data structure. Yet because a network model is called on to estimate the often very low likelihood values associated with long-distance events, it

is important to characterize the uncertainties associated with these estimates, since the uncertainties have some influence on their interpretation. In other words, when the estimated risk of invasion for a given location of interest is very low, it is important to understand how the estimate will behave given various uncertainties in the data and model assumptions, since they constrain the precision and thus reliability with which the estimate can be interpreted by a model user (Zeckhauser and Viscusi 1990).

In North America, there is increasing concern that forest pest invasions are being facilitated by the transport of firewood for recreational purposes (Haack et al. 2010; Tobin et al. 2010; Koch et al. 2012). The concern has been particularly motivated by the ongoing range expansion of the highly destructive emerald ash borer (*Agrilus planipennis* Fairmaire) in the eastern U.S. and Canada, as infested firewood is considered one of the insect's primary vectors of spread (Muirhead et al. 2006; Petrice and Haack 2006; Poland and McCullough 2006). Although there has been general research into the sociology behind recreational firewood usage in North America (see Jacobi et al. 2011; USDA Animal and Plant Health Inspection Service 2011), still little is known about geographic patterns of firewood movement by campers, and the implications of those patterns for invasions. In response, we developed a geographically explicit network model that assesses the likelihood of dispersal of forest pests in camper-transported firewood. We built the model using data from the U.S. National Recreation Reservation Service, which document visitor travel between the populated places where they reside and federal campgrounds across the U.S. and Canada. Among our initial objectives for this model was the ability to identify, for any location where a forest pest is believed to have been introduced via camper-transported firewood, the other location(s) from which the pest likely originated. Unfortunately, such results are not synoptic enough for broad-scale policy-making, which is rarely done at the level of individual urban areas. Therefore, we generated raster output maps that summarize, for individual U.S. states and Canadian provinces, where outside the target state or province an introduced forest pest is most likely to have originated.

Treating these output maps as a set of baseline scenarios, we explored the sensitivity of their "origin risk" estimates to errors in the probabilities of pest transmission between locations, as well as randomized removal of nodes from the underlying network. Here, we discuss the implications of these sources of uncertainty for the interpretation of the derived maps. The use of sensitivity analysis techniques to characterize the influence of uncertainty on model outputs is not unusual (Morgan and Henrion 1990; Helton and Davis 2002; Li and Wu 2006), although it remains relatively uncommon for spatially explicit analyses of invasive species risk (but see Venette and Cohen 2006; Koch et al. 2009). For this study, our primary objective was to demonstrate a simple way to identify aspects of our network model where the introduction of uncertainty provoked an inconsistent response across the set of output maps. This would allow us to determine whether any aspect of the model framework, rather than the data populating it, was a potentially problematic source of uncertainty. We view this as an essential first step in a broader analysis of uncertainty in invasive species risk maps and their underlying models.

Methods

Data source

We constructed our network model using a broad-scale data set from the U.S. National Recreation Reservation Service (NRRS), which operates an online reservation system for campgrounds and related facilities operated by the U.S. Forest Service, National Park Service, Bureau of Land Management, and other federal agencies. The available data spanned a period of greater than five years (January 2004-September 2009) and documented ~7.2 million visitor reservations, including reservations from Canada, at campgrounds and recreational facilities throughout the continental U.S. Although this is a large data set, there are many public and private campgrounds outside the NRRS system. Hence, we assumed the data to be a representative sample of all camper travel.

Each reservation record documented the visitor's origin location (i.e., by ZIP code or Canadian postal code), the destination campground, and the date of the visit. The initial processing of the NRRS data set is described in greater detail in Koch et al. (2012). Notably, the data did not indicate camper firewood usage, so our model instead depicts the general travel behaviour of all campers as documented in the data, a consistent proportion of whom we assumed were carrying potentially infested firewood (for additional discussion regarding this proportion, see Haack et al. 2010; Jacobi et al. 2011; Koch et al. 2012). We should also note that we ignored the visit date for this study, although we acknowledge that a visit's timing (e.g., in terms of a particular season) can affect the likelihood of pest emergence from firewood and subsequent establishment.

For each individual reservation, we calculated the geographic (Euclidean) distance between the visitor's origin location (i.e., the centroid of the polygon depicting the visitor's ZIP or postal code) and the destination campground. We then parsed the data into a set of unique pathway segments. Some data aggregation was necessary during this step to ensure model tractability. Conceptually, our network model is formulated as a first-order transition matrix (Karlin and Taylor 1975) that estimates the probability of travel between every possible origin-destination combination. Each reservation record in the NRRS data represented a single trip of some specified distance between a pair of origin and destination locations. Altogether, the data featured >50,000 visitor origin locations and >2500 destination locations, which translated to >973,000 pairwise combinations involving at least one visitor reservation (i.e., at least one trip) during the study period. To build a transition matrix using this many pairwise combinations was computationally impractical, so we aggregated the data at a coarse spatial resolution (into 16 × 16 km cells), ending up with a networked set of ~14,000 unique map units. These cells provided complete spatial coverage of our study area, which included both the continental U.S. and most of Canada.

Model structure

Any two map cells, designated i and j, in this networked set were connected by a unique pathway segment, ij. Each ij had a value, m_{ij}, representing the number of trips from i to j documented in the aggregated NRRS data during the study time period t (i.e., January 2004-September 2009), and another value, m_{ji}, representing the number of return trips from j back to i. We constructed a matrix, **M**, from these m_{ij} and m_{ji} values. The size of the matrix was $n \times n$, corresponding to the number of unique map units in the aggregated data (i.e., n = ~14,000):

$$\mathbf{M} = \begin{bmatrix} 0 & m_{12} & \cdots & m_{1n} \\ m_{21} & 0 & \cdots & m_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ m_{n1} & m_{n2} & \cdots & 0 \end{bmatrix} \quad (1)$$

The non-diagonal elements of **M** were the m_{ij} and m_{ji} values for the pathway segments connecting each (i, j) pair of map cells. The diagonal elements of **M**, where i = j, were set to 0.

Based on **M**, we constructed a probability matrix, $\mathbf{P}_t$, of camper travel (and by extension, transport of potentially infested firewood) along the pathway segments during the study period t. We assumed that the probability, p_{ij}, of travel along any given segment ij was linearly related to the number of trips from i to j during period t:

$$p_{ij} = m_{ij}\lambda_t, \quad (2)$$

where λ_t is a scaling parameter. Ideally, the parameter λ_t would define the total likelihood (i.e., over the study period t) of camper transport of pest-infested firewood from i to j in the pathway matrix. In fact, a precise value of λ_t would be necessary to estimate an exact probability of a pest being moved from location i to j. Our objective, however, was not to derive exactly estimated output probabilities, but the more conservative goal of testing the relative sensitivity of model outputs to changes in the inputs, such that the output probabilities simply serve as a relative measure of risk. For this purpose, we only needed λ_t to be sufficiently small to ensure that the sum of the p_{ij} probabilities in each $\mathbf{P}_t$ matrix row was below 1.

Each row of $\mathbf{P}_t$ included another variable, $p_{i\,term}$, representing the probability that no camper travel (i.e., to any j) would originate at i:

$$p_{i\,term} = 1 - \sum_{j=1}^{n} p_{ij} \quad (3)$$

If $p_{i\,term}$ was equal to 1 for any row, then the map cell i associated with that row did not serve as a point of origin in the model. This did not preclude the cell from serving as a potential destination j. The probability matrix $\mathbf{P}_t$ of camper travel along each pathway segment (i.e., during period t) was thus specified as follows:

$$\mathbf{P}_t = \lambda_t \mathbf{M} = \begin{bmatrix} 0 & p_{12} & \cdots & p_{1n} & p_{1term} \\ p_{21} & 0 & \cdots & p_{2n} & p_{2term} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ p_{n1} & p_{n2} & \cdots & 0 & p_{nterm} \end{bmatrix} \quad (4)$$

We used $\mathbf{P}_t$ as the basis for 2×10^6 stochastic pathway simulations of camper travel from each cell i to other map cells. Setting i as the origin location, the model extracted the vector of travel probabilities associated with i (i.e., the probabilities in row i) from $\mathbf{P}_t$ in order to simulate subsequent camper travel from i through the pathway network to other map cells. Briefly, in each simulation run for a given cell i, the model performed a uniform random draw against the associated row of probabilities in $\mathbf{P}_t$ in order to select the next map cell defining the path of an individual trip. The simulation of this path continued until reaching a final destination map cell (i.e., a cell with no outgoing travel), or instead, when a terminal state (i.e., no further travel) was selected based on the $p_{i\,term}$ value. Regardless of the number of individual segments comprising a simulated path, we assumed that the path was completely traversed within the study period t. Consequently, for a given pathway segment ij, the probability of camper travel from location i to location j during period t was estimated as follows:

$$\varphi_{ij} = M_{ij}/M, \quad (5)$$

where M_{ij} is the number of times pathway travel from location i to j was simulated to occur, and M is the total number of simulations ($M = 2 \times 10^6$ in this study).

Out-of-state origin risk maps

The network model permitted us to generate maps where each individual cell in the study area could be set as the origin or destination location of interest. While cell-specific maps might have value when addressing specific invasion scenarios (e.g., identifying the probable source locations for a single cell found to be invaded), they do not provide a comprehensive picture of invasion risk for use in setting management priorities. So, we chose to summarize the model results in a broader context. In a previous analysis of the NRRS data (Koch et al. 2012), we found that the majority (~53%) of camping trips involved travel distances of less than 100 km. However, we also found that 10% of trips involved distances of greater than 500 km. This is consistent with other research (Haack et al. 2010; Jacobi et al. 2011; USDA Animal and Plant Health Inspection Service 2011)

suggesting that campers frequently transport firewood across state or provincial borders. Furthermore, much of the regulatory decision-making with respect to invasive species, including the implementation of firewood restrictions to limit forest pest spread, happens at the state or provincial level (Filbey et al. 2002). For these reasons, we opted to develop model outputs that characterized those locations (i.e., map cells) outside a state or province of interest that were most strongly linked to locations inside the state or province.

For each U.S. state and Canadian province, we generated a map where, for each cell i outside the target state (province), we summed the model-derived travel probabilities (i.e., the φ_{ij} values) for all pathways between i and any destination cell j within the target state (province). Essentially, each map depicts the most likely external source locations if a firewood-transported forest pest were to be discovered within the state (province) of interest. (Note that for the Canadian provinces, we assume that campers returning from U.S. locations may be transporting forest pests, despite border biosecurity measures.) A principal feature of these maps is that they highlight key "bottleneck" locations where surveillance, awareness campaigns, or quarantine procedures have the best potential to be cost-effective in terms of protecting the state (province) of interest from invasion (Hauser and McCarthy 2009). These maps served as a set of baseline maps that we could then compare to the maps generated for a set of sensitivity scenarios, described below.

Sensitivity analyses

We performed sensitivity analyses to evaluate how uncertainty in key aspects of the network model's structure would influence the φ_{ij} probabilities estimated in the baseline maps. The primary goal of these analyses was to assess whether uncertainties in any of the tested model aspects yielded a set of output maps with spatial patterns that departed in unexpected ways from the patterns depicted in the corresponding baseline maps. We adopted a Monte Carlo simulation approach (Morgan and Henrion 1990; Crosetto and Tarantola 2001; Li and Wu 2006) for the analyses, involving three basic steps: random sampling from an input distribution defined for each aspect of interest; repeated model runs using the randomly sampled values; and summarization of these results for comparison to the baseline maps.

We focused on three model aspects for this study. Our first sensitivity scenario tested the impact of uncertainty in the network configuration by randomly removing up to 30% of nodes (i.e., map cells and their associated vectors of p_{ij} probabilities) from the network prior to each new model run. Our second sensitivity scenario tested the impact of uncertainty in the scaling parameter λ_t (see Eq. 2). In this case, we added uniform random error within the range ±0.15 to λ_t prior to completing each new model run. Our final sensitivity scenario evaluated the impact of uncertainty in the p_{ij} values comprising $\mathbf{P}_t$. To test this aspect, we added uniform random error of up to 0.01 to the p_{ij} values before each new model run. Ideally, we would have repeated the three scenarios with series of bounding values. However, because of the computational complexity of this exercise, we chose a single bounding value for each sensitivity scenario

that, based on preliminary work with the model, we were confident would provide meaningful contrast to the baseline scenario.

We completed 2×10^6 pathway simulations for each sensitivity scenario. This process yielded scenario-specific maps of the φ_{ij} probabilities (see Eq. 5) for each state and province that we could compare to the corresponding baseline maps. For this comparison, we calculated and mapped the differences, in percentage terms, between the output probabilities under the baseline scenario and under each sensitivity scenario. We also calculated a Moran's *I* (Moran 1950) value for each difference map. Moran's *I* is a measure of the global (i.e., map-wide) spatial autocorrelation of values in a spatially referenced data set (Getis and Ord 1992; Perry et al. 2002). It ranges in value from -1 to 1, with values close to -1 indicating a high degree of dispersion in the data (e.g., a checkerboard spatial pattern), values close to 1 indicating a high degree of clustering, and values near 0 indicating a random spatial pattern.

We employed a simple rank-difference approach for further comparison of the baseline results to those from our third sensitivity scenario dealing with uncertainty in the p_{ij} values. After calculating the per-cell differences between the baseline and sensitivity maps for each state and province, we converted both the baseline probabilities and the calculated differences into ranks. Each cell in a given map was assigned a rank from 1 to 6 based on a global percentile distribution of the values that occurred in the maps for all states and provinces; cell values that fell within the bottom 25% of this global distribution received a rank of 1, cells in the 25–75% range received a rank of 2, cells in the 75–90% range received a rank of 3, cells in the 90–95% range received a rank of 4, cells in the 95–99% range received a rank of 5, and cells in the top 1% of this distribution received a rank of 6. We ranked each cell twice in this fashion, first according to its percentile value under the baseline scenario, then according to its percentile value in the global distribution of differences under this sensitivity scenario. Next, we calculated the change in rank between the two, which yielded an index value ranging from -5 to 5. Finally, we mapped the index values for each state and province in order to identify any spatial trends.

Results

Baseline scenario

Across all individual state and province maps created for the baseline scenario, we saw three general spatial patterns of out-of-state (or out-of-province) origin risk. First, as illustrated by the state of Alabama (Fig. 1a), there were cases where the highest origin probabilities (i.e., the per-cell φ_{ij} values) were primarily limited to a narrow fringe zone surrounding the target state. This pattern was common among states in the southeastern and northeastern U.S. The second general pattern, exemplified by Missouri (Fig. 1b), similarly featured a localized zone of high probabilities around the target state, but this was augmented by several high-probability hotspots associated with major urban areas located outside this zone. For instance, the map for Missouri

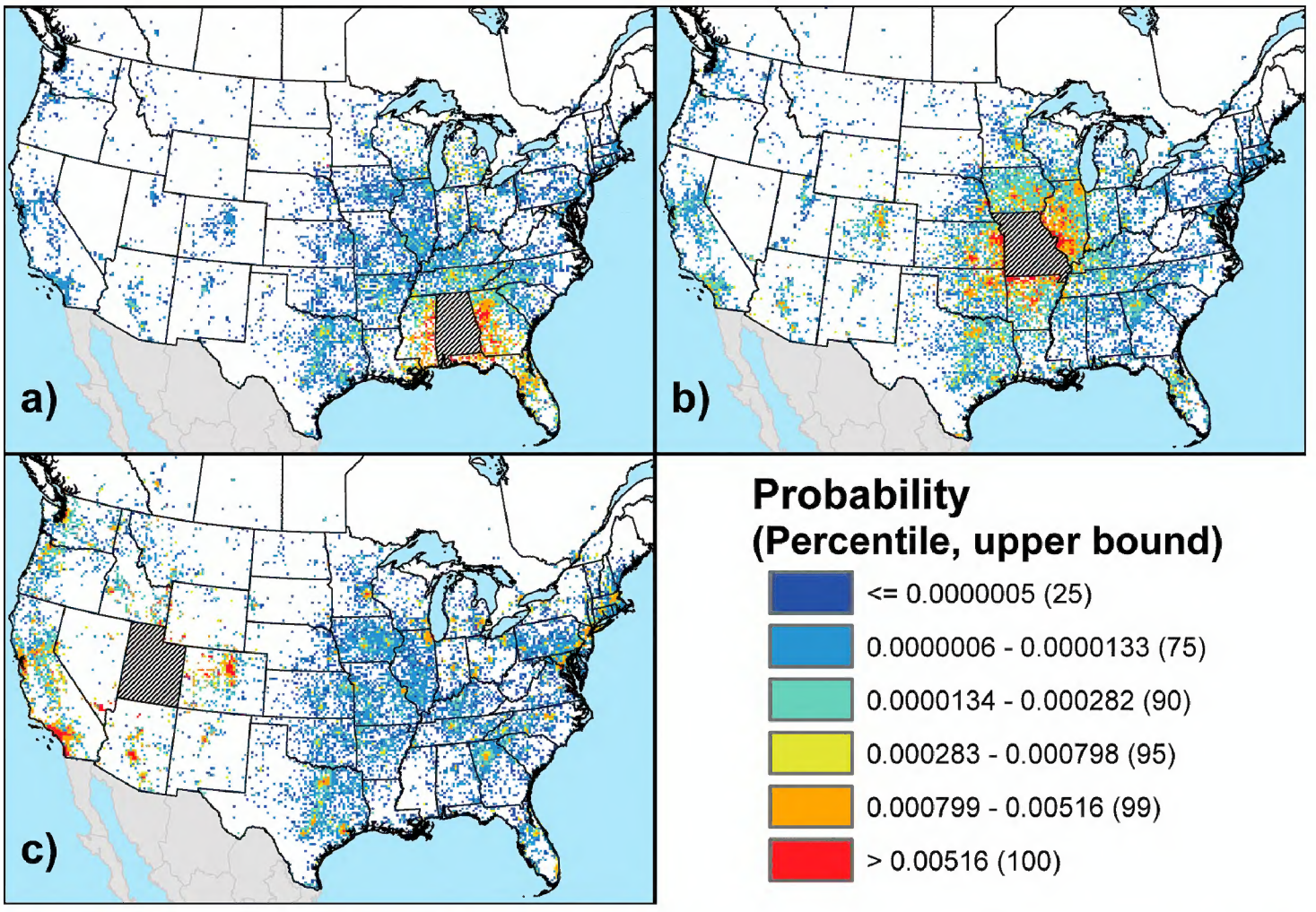

Figure 1. Out-of-state origin risk maps for three U.S. states under the baseline scenario: **a** Alabama **b** Missouri **c** Utah. The maps show the relative probability, φ_{ij}, that a map cell outside the state of interest is a source of campers carrying potentially infested firewood to the target state. Numbers in parentheses link the defined probability classes to a global percentile distribution built from the probabilities that occurred in the maps for all states and provinces.

included hotspots associated with the cities of Chicago (IL), Dallas (TX), Denver (CO), all of which are at least 400 km away. This mixed origin pattern was typical of many states in the U.S. Midwest. In the third pattern, exemplified by Utah (Fig. 1c), there was also a subtle local fringe zone, but the majority of map cells with high probabilities were associated with major urban areas that are fairly distant from the target state. With respect to Utah, the most prominent urban areas (i.e., those displaying uniformly high probabilities) were generally in the western U.S., but a number of eastern U.S. cities also had probabilities in the top 5% of the global distribution of probabilities across all states and provinces. This sort of dispersed pattern of origin risk was exhibited by other states such as Arizona and California that, like Utah, feature several popular recreational destinations (e.g., national parks) that draw campers from throughout the continent.

The maps for the Canadian provinces, especially the more populous ones (i.e., Quebec, Ontario, Alberta, and British Columbia), usually exhibited a dispersed spatial pattern similar to that observed for Utah, although the origin probabilities in each map were low compared to their U.S. counterparts. Indeed, the highest probabilities in these maps very rarely reached the top 10% of the global distribution. The only exceptions were a few high-probability (top 5%) cells in each of the maps for the most popu-

lous Canadian provinces, which appeared to be associated with specific recreational destinations such as Grand Canyon National Park (AZ) and Zion National Park (UT).

Two factors seem to govern the observed patterns of origin risk. Foremost is the population of the state or province of interest. Because the model is bi-directional (i.e., it also simulates return travel by campers from their destinations to their origin locations), the population of the target state or province influences the range of output probabilities that will be portrayed in its map, particularly the frequency at which high probabilities occur. In short, the maps of heavily populated states and provinces tended to have higher probabilities overall than sparsely populated states and provinces. The second factor is the nature of the recreational opportunities available in the target state or province. States or provinces containing several high-profile recreational destinations displayed a dispersed pattern of origin risk, while states or provinces with numerous but comparatively low-profile recreational destinations usually displayed a more localized origin risk pattern.

Impact of uncertainty in the network configuration

Figure 2 shows three example maps of the mean percent difference (i.e., over all model runs) in the φ_{ij} output probabilities under this sensitivity scenario. The example states

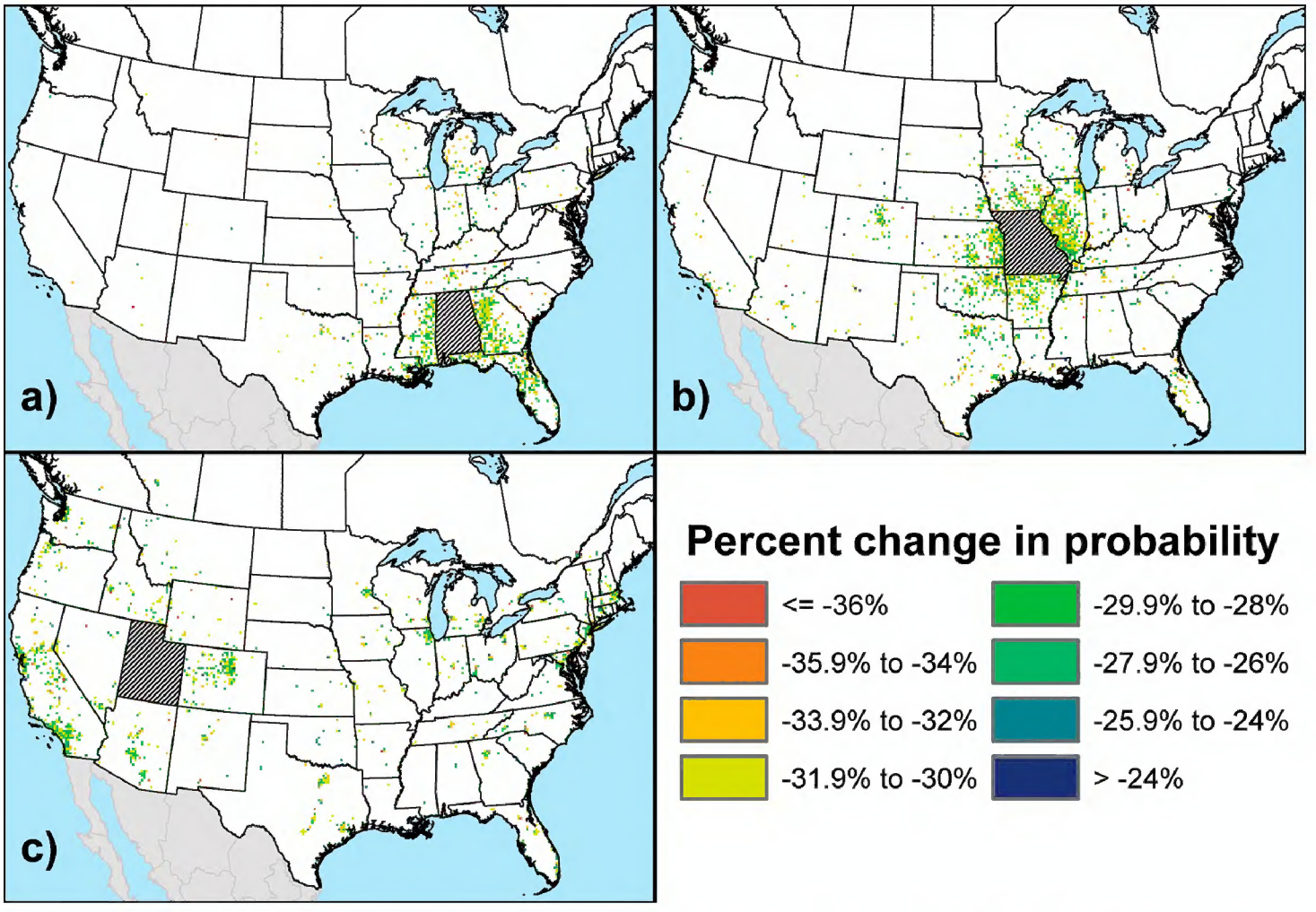

Figure 2. Maps of the mean percent difference (i.e., across all model runs) between the φ_{ij} output probabilities calculated under the baseline scenario and under the sensitivity scenario that tested the impact of uncertainty in the network configuration: **a** Alabama **b** Missouri **c** Utah. Only cells with probabilities in the top 10% of the global distribution for the baseline scenario are shown.

correspond to those highlighted under the baseline scenario: Alabama (Fig. 2a), Missouri (Fig. 2b), and Utah (Fig. 2c). Note that only map cells in the top 10% of probabilities under the baseline scenario are shown. For all states and provinces, the percent change in the values of cells in this top 10% typically ranged from -24% to -36%, with a map-wide mean (i.e., across all cells in a given map) near -30%. This level of difference is consistent with the proportion of nodes removed for this scenario. More importantly, as illustrated by the difference maps for all three example states, there was no obvious spatial trend in the pattern of change. Essentially, each map has a salt-and-pepper appearance, such that cells with higher probabilities under the baseline scenario (see Fig. 1) do not appear to have greater (or smaller) percent changes than cells with lower probabilities under the baseline. This observation is supported by the Moran's *I* values calculated for all states and provinces, which ranged between -0.016 and 0.021 (median = 0.005) and thus indicated a minimal level of spatial autocorrelation.

Impact of uncertainty in the scaling parameter λ_t

Figure 3 shows three example maps of the mean percent difference in the φ_{ij} output probabilities under this sensitivity scenario. As in the previous scenario, the example

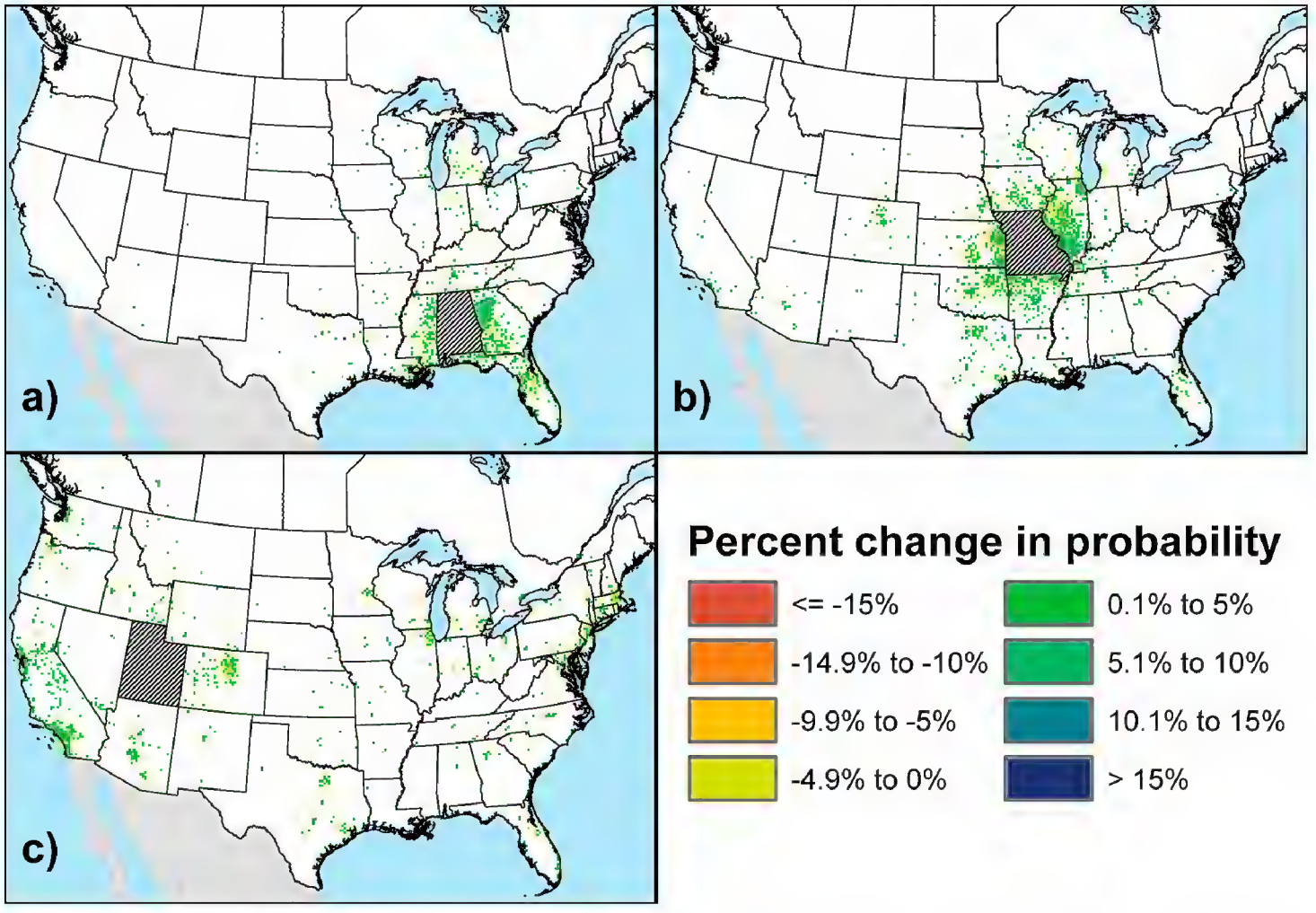

Figure 3. Maps of the mean percent difference (i.e., across all model runs) between the φ_{ij} output probabilities calculated under the baseline scenario and under the sensitivity scenario that tested the impact of uncertainty in the scaling parameter λ_t: **a** Alabama **b** Missouri **c** Utah. Only cells with probabilities in the top 10% of the global distribution for the baseline scenario are shown.

states correspond to those highlighted under the baseline scenario: Alabama (Fig. 3a), Missouri (Fig. 3b), and Utah (Fig. 3c). Only map cells in the top 10% of probabilities under the baseline scenario are shown. For all states and provinces, the percent change in the values of cells in this top 10% typically ranged from -15% to 15%, with a map-wide mean slightly greater than 0. This degree of change is consistent with the alterations to λ_t under this scenario.

None of the difference maps for the example states appears to show a strong spatial trend in the pattern of change. There may be a slight tendency for map cells with higher probabilities under the baseline scenario to exhibit small but positive percent changes, while cells with lower probabilities under the baseline may tend toward small negative changes. However, the existence of a subtle spatial trend is not really supported by the Moran's *I* values for each state and province, which ranged from -0.005 to 0.133 (median=0.008). Indeed, with the exception of a single Canadian province, Nova Scotia, which has relatively few connections to outside locations in our network model, none of the maps had an *I* value greater than 0.0324. Similar to the previous sensitivity scenario, this suggests a low level of spatial autocorrelation.

Impact of uncertainty in the p_{ij} probabilities

Figure 4 shows three example maps of the rank differences (see Methods) between the baseline scenario and this sensitivity scenario, which tested the effects of adding uniform random error to the p_{ij} probabilities. The maps display an obvious spatial trend in the pattern of change. (Again, only cells in the top 10% of probabilities under the baseline scenario are shown.) This trend can be summarized in general terms: cells with comparatively higher probabilities under the baseline scenario were more likely to exhibit a decrease or no change in rank under this sensitivity scenario, while cells with comparatively lower probabilities under the baseline were more likely to exhibit an increase in rank. Thus, for states and provinces like Alabama (Fig. 4a), where the highest origin probabilities under the baseline scenario were primarily limited to a localized zone adjoining the state or province (also see Fig. 1a), any decreases in rank were also usually limited to this zone. Alternatively, in states and provinces like Missouri (Fig. 4b) and Utah (Fig. 4c), where high origin probabilities under the baseline were regularly associated with major urban areas outside the fringe zone, it is primarily map cells in these areas that exhibited a decrease or, just as commonly, no change in rank. For all states and provinces, the map cells that exhibited increases in rank under this sensitivity scenario were usually found in rural areas or the peri-urban zones (Allen 2003) that surround large cities.

The existence of a spatial trend in the pattern of change is supported by the Moran's *I* values calculated for the difference maps of each state and province. With the exception of Nova Scotia, the values ranged from 0.057 to 0.208 (median = 0.135), indicating a degree of positive spatial autocorrelation (i.e., clustering) that varied according to whether the state or province of interest had a well-defined fringe zone of

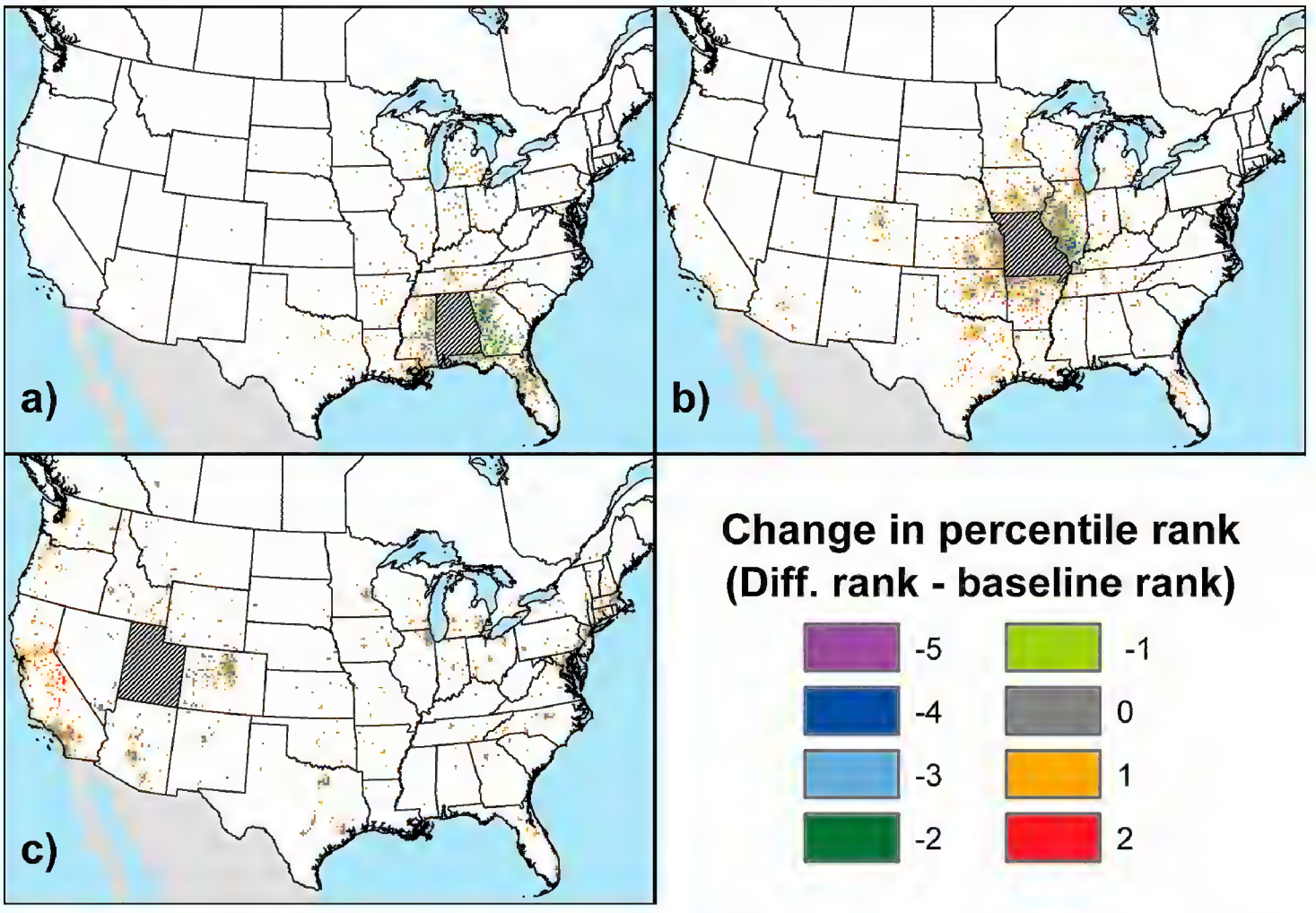

Figure 4. Maps showing the difference in rank between maps generated under the baseline scenario and under the sensitivity scenario that tested the impact of uncertainty in the p_{ij} probabilities: **a** Alabama **b** Missouri **c** Utah. See main text for details regarding the calculation of ranks, which ranged from 1 to 6 for each scenario. Only cells with probabilities in the top 10% of the global distribution for the baseline scenario are shown.

high probabilities under the baseline scenario. Typically, states and provinces with such a zone exhibited larger I values. This variability, however, does not contradict the notion of a generally consistent spatial trend across all states and provinces.

Discussion

By focusing on changes in the φ_{ij} probabilities under the sensitivity scenarios, we can make conclusions about how the model generally behaves given uncertainty in the tested model aspects. We expected to see departures from the baseline φ_{ij} values in all three scenarios, but importantly, our results indicated that the patterns of these differences were consistent across all states and provinces. In the two scenarios testing the impact of uncertainty in the network configuration and in the scaling parameter λ_t, the distribution of the mapped differences was usually close to random; in other words, no particular state (province) or range of output probabilities was uniquely affected by uncertainty in these model aspects. In the scenario that tested the impact of uncertainty in the p_{ij} values, there was a recognizable spatial trend in the differences from the baseline scenario, but this trend was manifested similarly across all states and

provinces. Essentially, uncertainty in the p_{ij} values resulted in smoothing of the φ_{ij} output probabilities across the entire map area, regardless of how those probabilities were spatially distributed under the baseline scenario.

In summary, it appears that the network model is generally stable in the presence of uncertainties in its critical structural aspects (i.e., in its spatial configuration and in parameters that define $\mathbf{P}_t$, the probability matrix that drives the pathway simulations). While making this determination was the primary objective of our study, close examination of the individual sensitivity scenarios reveals additional details about potential model improvements and future applications. First, removing a large portion of the nodes from the network did not cause the model to behave inconsistently. This stability suggests that the modelling framework may be transferable to other data sets that can be organized as networks, including potentially sparser data sets. This transferability may present an easy opportunity to examine other modes of human-mediated dispersal that are relevant to invasion risk. For instance, we might want to apply the model to describe camper travel patterns associated with privately owned campgrounds across North America. The framework may also be applicable to the movement of crop pests via domestic shipments of certain agricultural commodities. Similarly, the network model continued to behave consistently despite sizeable alterations to the scaling parameter λ_t. This appears to affirm our supposition that a more precise value for λ_t would not substantially affect the generalbehaviour of the model, which further supports the notion of its transferability to other invasion risk modelling problems.

In the sensitivity scenario that tested the impact of uncertainty in the p_{ij} probabilities, the smoothing effect observed in the output maps suggests that these probabilities were fairly sensitive to uncertainty, or at least more sensitive than the other tested model aspects. Under this scenario, a small random variate was added to every p_{ij} value in the $\mathbf{P}_t$ matrix, including where $p_{ij} = 0$. This change effectively created new interconnecting vectors between network nodes, subsequently adding topological information to the network. Indirectly, this raised the relative importance of map cells with low origin probabilities under the baseline scenario, since it added some small probability of camper travel between cells, even in cases where no such travel was documented in the NRRS data.

Conceptually, this sensitivity scenario depicts a general lack of knowledge about the travel patterns of campers (and by extension, their movement of potentially infested firewood). This implies that the best opportunity to improve our model may be to refine the p_{ij} probabilities. Notably, these values are partly shaped by the parameter λ_t (see Eq. 2), so the issues mentioned previously regarding λ_t may have relevance here. Indeed, while we did not implement a sensitivity scenario where λ_t and the p_{ij} values were varied together, it is possible that the p_{ij} values are more or less sensitive depending on the value of λ_t. This is a potential topic of future work. Regardless, the p_{ij} values are directly derived from the data underlying the network, and so depend on the comprehensiveness and representativeness of those data for the phenomenon being modelled. In our case, the NRRS data only cover a limited subset of all camper travel, so there could be some advantage to including other data sources such as camper res-

ervation records at private- or state/provincial-managed campgrounds, contingent on the availability of such data.

We must reiterate that the φ_{ij} output probabilities only provide a relative measure of invasion risk. Because we did not have sufficient data to directly model the movement of forest pests in infested firewood, and instead modelled the travel of all campers as a proxy, then the resulting probabilities should not be interpreted as a depiction of the true risk. The actual probabilities of successful forest pest introduction via firewood are likely much lower than the numbers we present here, for various reasons; for instance, not all campers transport firewood, not all of the firewood they transport is infested, and not all infested firewood contains enough individuals of a potential pest species to support establishment upon arriving at a destination (Haack et al. 2010; Haack et al. 2011; Jacobi et al. 2011; Koch et al. 2012). Being able to better define λ_t might bring us closer to estimating the true probabilities. However, it may be worth considering whether this would be a cost-effective expenditure of research effort. For example, developing a continuous function relating travel distance to the likelihood a camper is transporting firewood might require extensive surveys of camper behaviour, and then would only provide a partial estimate of λ_t. Moreover, a relative measure of invasion risk may still be quite pertinent for a common use of pest risk maps and models (Venette et al. 2010; Yemshanov et al. 2012): to prioritize locations for surveillance or other pest management activities.

While we have proposed that network models may be especially useful for mapping pest risk, they, like other types of models, should be subject to critical scrutiny prior to their application. We see the presented approach as being primarily of interest to analysts tasked with constructing network models. The approach does not address key questions that arise in uncertainty analysis regarding how uncertainty propagates to the model outputs, or the resulting implications for decision makers who will utilize these outputs. Instead, it is merely intended to help analysts assess the fundamental soundness of their models and feel more confident about incorporating them into their analytical toolsets.

Conclusions

Because human activities contribute significantly to the proliferation of invasive species, modelling approaches that characterize important human-mediated dispersal pathways may be highly applicable for pest risk analysis. To provide a working example, we developed a geographically explicit network model that depicts the potential spread of forest pests in firewood moved by camper travel across the U.S. and Canada. We then presented an approach, based on common sensitivity analysis techniques, for assessing how this network model behaves when uncertainty is introduced into critical model aspects. In our case, the approach allowed us to determine that the model behaved consistently and predictably in the presence of uncertainty. The approach is analytically straightforward, and should be generalisable to other network models with comparable formulations.

Acknowledgements

We thank Judith Pasek (USDA APHIS, retired) and David Kowalski (USDA APHIS) for providing the NRRS data, as well as Roger Magarey (North Carolina State University) and Daniel Borchert (USDA APHIS) for their assistance with the project. The participation of Denys Yemshanov was supported by an interdepartmental NRCan - CFIA Forest Invasive Alien Species initiative.

References

Allen A (2003) Environmental planning and management of the peri-urban interface: perspectives on an emerging field. Environment and Urbanization 15: 135–148. doi: 10.1177/095624780301500103

Andow DA, Kareiva PM, Levin SA, Okubo A (1990) Spread of invading organisms. Landscape Ecology 4: 177–188. doi: 10.1007/BF00132860

Carlton JT, Ruiz GM (Eds) (2003) Invasive species: vectors and management strategies. Island Press, Washington, D.C., 484 pp.

Crosetto M, Tarantola S (2001) Uncertainty and sensitivity analysis: tools for GIS-based model implementation. International Journal of Geographical Information Science 15: 415–437. doi: 10.1080/13658810110053125

Filbey M, Kennedy C, Wilkinson J, Balch J (2002) Halting the invasion: state tools for invasive species management. Environmental Law Institute, Washington, D.C., 112 pp.

Floerl O, Inglis GJ, Dey K, Smith A (2009) The importance of transport hubs in stepping-stone invasions. Journal of Applied Ecology 46: 37–45. doi: 10.1111/j.1365-2664.2008.01540.x

Getis A, Ord JK (1992) The analysis of spatial association by use of distance statistics. Geographical Analysis 24: 189–206. doi: 10.1111/j.1538-4632.1992.tb00261.x

Haack RA, Petrice TR, Wiedenhoft AC (2010) Incidence of bark- and wood-boring insects in firewood: a survey at Michigan's Mackinac Bridge. Journal of Economic Entomology 103: 1682–1692. doi: 10.1603/EC10041

Haack RA, Uzunovic A, Hoover K, Cook JA (2011) Seeking alternatives to probit 9 when developing treatments for wood packaging materials under ISPM No. 15. EPPO Bulletin 41: 39–45. doi: 10.1111/j.1365-2338.2010.02432.x

Harwood TD, Tomlinson I, Potter CA, Knight JD (2010) Dutch elm disease revisited: past, present and future management in Great Britain. Plant Pathology 220: 3353–3361. doi: 10.1111/j.1365-3059.2010.02391.x

Hastings A (1996) Models of spatial spread: a synthesis. Biological Conservation 78: 143–148. doi: 10.1016/0006-3207(96)00023-7

Hastings A, Cuddington K, Davies KF, Dugaw CJ, Elmendorf S, Freestone A, Harrison S, Holland M, Lambrinos J, Malvadkar U, Melbourne BA, Moore K, Taylor C, Thomson D (2005) The spatial spread of invasions: new developments in theory and evidence. Ecology Letters 8: 91–101. doi: 10.1111/j.1461-0248.2004.00687.x

Hauser CE, McCarthy MA (2009) Streamlining 'search and destroy': cost-effective surveillance for invasive species management. Ecology Letters 12: 683–692. doi: 10.1111/j.1461-0248.2009.01323.x

Helton JC, Davis FJ (2002) Illustration of sampling-based methods for uncertainty and sensitivity analysis. Risk Analysis 22: 591–622. doi: 10.1111/0272-4332.00041

Horan RD, Perrings C, Lupi F, Bulte EH (2002) Biological pollution prevention strategies under ignorance: the case of invasive species. American Journal of Agricultural Economics 84: 1303–1310. doi: 10.1111/1467-8276.00394

Hulme PE (2009) Trade, transport and trouble: managing invasive species pathways in an era of globalization. Journal of Applied Ecology 46: 10–18. doi: 10.1111/j.1365-2664.2008.01600.x

Hulme PE, Bacher S, Kenis M, Klotz S, Kühn I, Minchin D, Nentwig W, Olenin S, Panov V, Pergl J, Pyšek P, Roques A, Sol D, Solarz W, Vilà M (2008) Grasping at the routes of biological invasions: a framework for integrating pathways into policy. Journal of Applied Ecology 45: 403–414. doi: 10.1111/j.1365-2664.2007.01442.x

Jacobi WR, Goodrich BA, Cleaver CM (2011) Firewood transport by national and state park campers: a risk for native or exotic tree pest movement. Arboriculture & Urban Forestry 37: 126–138

Kaluza P, Kölzsch A, Gastner MT, Blasius B (2010) The complex network of global cargo ship movements. Journal of the Royal Society Interface 7: 1093–1103. doi: 10.1098/rsif.2009.0495

Karlin S, Taylor HM (1975) A first course in stochastic processes. Academic Press, New York, 557 pp.

Keller RP, Frang K, Lodge DM (2008) Preventing the spread of invasive species: economic benefits of intervention guided by ecological predictions. Conservation Biology 22: 80–88. doi: 10.1111/j.1523-1739.2007.00811.x

Koch FH, Yemshanov D, Magarey RD, Smith WD (2012) Dispersal of invasive forest insects via recreational firewood: a quantitative analysis. Journal of Economic Entomology 105: 438–450. doi: 10.1603/EC11270

Koch FH, Yemshanov D, McKenney DW, Smith WD (2009) Evaluating critical uncertainty thresholds in a spatial model of forest pest invasion risk. Risk Analysis 29: 1227–1241. doi: 10.1111/j.1539-6924.2009.01251.x

Lewis MA, Pacala S (2000) Modeling and analysis of stochastic invasion processes. Journal of Mathematical Biology 41: 387–429. doi: 10.1007/s002850000050

Li H, Wu J (2006) Uncertainty analysis in ecological studies: an overview. In: Wu J, Jones KB, Li H, Loucks OL (Eds) Scaling and uncertainty analysis in ecology: methods and applications. Springer, Dordrecht, The Netherlands, 43–64.

Moran PAP (1950) Notes on continuous stochastic phenomena. Biometrika 37: 17–23. doi: 10.1093/biomet/37.1-2.17

Morgan MG, Henrion M (1990) Uncertainty: a guide to dealing with uncertainty in quantitative risk and policy analysis. Cambridge University Press, New York, NY, 332 pp. doi: 10.1017/CBO9780511840609

Moslonka-Lefebvre M, Finley A, Dorigatti I, Dehnen-Schmutz K, Harwood T, Jeger MJ, Xu X, Holdenrieder O, Pautasso M (2011) Networks in plant epidemiology: from genes to landscapes, countries, and continents. Phytopathology 101: 392–403. doi: 10.1094/phyto-07-10-0192

Muirhead JR, Leung B, van Overdijk C, Kelly DW, Nandakumar K, Marchant KR, MacIsaac HJ (2006) Modelling local and long-distance dispersal of invasive emerald ash borer *Agrilus planipennis* (Coleoptera) in North America. Diversity and Distributions 12: 71–79. doi: 10.1111/j.1366-9516.2006.00218.x

Nathan R (2006) Long-distance dispersal of plants. Science 313: 786–788. doi: 10.1126/science.1124975

Nathan R, Perry G, Cronin JT, Strand AE, Cain ML (2003) Methods for estimating long-distance dispersal. Oikos 103: 261–273. doi: 10.1034/j.1600-0706.2003.12146.x

Perry JN, Liebhold AM, Rosenberg MS, Dungan J, Miriti M, Jakomulska A, Citron-Pousty S (2002) Illustrations and guidelines for selecting statistical methods for quantifying spatial pattern in ecological data. Ecography 25: 578–600. doi: 10.1034/j.1600-0587.2002.250507.x

Petrice TR, Haack RA (2006) Effects of cutting date, outdoor storage conditions, and splitting on survival of *Agrilus planipennis* (Coleoptera: Buprestidae) in firewood logs. Journal of Economic Entomology 99: 790–796. doi: 10.1603/0022-0493-99.3.790

Poland TM, McCullough DG (2006) Emerald ash borer: invasion of the urban forest and the threat to North America's ash resource. Journal of Forestry 104: 118–124

Sakai AK, Allendorf FW, Holt JS, Lodge DM, Molofsky J, With KA, Baughman S, Cabin RJ, Cohen JE, Ellstrand NC, McCauley DE, O'Neil P, Parker IM, Thompson JN, Weller SG (2001) The population biology of invasive species. Annual Review of Ecology and Systematics 32: 305–332. doi: 10.1146/annurev.ecolsys.32.081501.114037

Sharov AA, Liebhold AM (1998) Model of slowing the spread of gypsy moth (Lepidoptera: Lymantriidae) with a barrier zone. Ecological Applications 8: 1170–1179. doi: 10.1890/1051-0761(1998)008[1170:MOSTSO]2.0.CO;2

Tobin PC, Diss-Torrance A, Blackburn LM, Brown BD (2010) What does "local" firewood buy you? Managing the risk of invasive species introduction. Journal of Economic Entomology 103: 1569–1576. doi: 10.1603/EC10140

USDA Animal and Plant Health Inspection Service (2011) Risk assessment of the movement of firewood within the United States. U.S. Department of Agriculture, Animal and Plant Health Inspection Service, Plant Protection and Quarantine, Center for Plant Health Science and Technology, Plant Epidemiology and Risk Analysis Laboratory, Raleigh, NC, 122 pp.

Venette RC, Cohen SD (2006) Potential climatic suitability for establishment of *Phytophthora ramorum* within the contiguous United States. Forest Ecology and Management 231: 18–26. doi: 10.1016/j.foreco.2006.04.036

Venette RC, Kriticos DJ, Magarey RD, Koch FH, Baker RHA, Worner SP, Raboteaux NNG, McKenney DW, Dobesberger EJ, Yemshanov D, Barro PJD, Hutchison WD, Fowler G, Kalaris TM, Pedlar J (2010) Pest risk maps for invasive alien species: a roadmap for improvement. BioScience 60: 349–362. doi: 10.1525/bio.2010.60.5.5

Wilson JRU, Dormontt EE, Prentis PJ, Lowe AJ, Richardson DM (2009) Something in the way you move: dispersal pathways affect invasion success. Trends in Ecology & Evolution 24: 136–144. doi: 10.1016/j.tree.2008.10.007

Yemshanov D, Koch FH, Lyons DB, Ducey M, Koehler K (2012) A dominance-based approach to map risks of ecological invasions in the presence of severe uncertainty. Diversity and Distributions 18: 33–46. doi: 10.1111/j.1472-4642.2011.00848.x

Zeckhauser RJ, Viscusi WK (1990) Risk within reason. Science 248: 559–564. doi: 10.1126/science.2333509